SR-L0-00456
JSC-16538
JUL 14 1980

8.0-10.3 0.3
NASA CR-
160754

AgRISTARS

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

April 1980

EVALUATING THE USE OF ANALYST LABELS IN MAXIMUM LIKELIHOOD CLUSTER PROPORTION ESTIMATION

R. K. Lennington and G. R. Terrell

(E80-10303) EVALUATING THE USE OF ANALYST LABELS IN MAXIMUM LIKELIHOOD CLUSTER PROPORTION ESTIMATION (Lockheed Engineering and Management) 14 p HC A02/MF A01 CSCL 05B
N80-30870
Unclas
G3/43 00303

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.
1830 NASA Road 1, Houston, Texas 77058

SR- LO-00456
JSC-16538

EVALUATING THE USE OF ANALYST LABELS IN MAXIMUM LIKELIHOOD CLUSTER PROPORTION ESTIMATION

Job Order 73-302

This report describes Classification activities of the Supporting Research project of the AgRISTARS program.

PREPARED BY

R. K. Lennington and G. R. Terrell

APPROVED BY

T. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

J. E. Wainwright

J. E. Wainwright, Manager
Development and Evaluation Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division

Space and Life Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

April 1980

LEMSCO-14672

PREFACE

The research which is the subject of this report was performed by personnel of the Lockheed Engineering and Management Services Company for the Supporting Research project of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing program.

This program is underway within the Earth Observations Division, Space and Life Sciences Directorate, at the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, Texas.

CONTENTS

1. INTRODUCTION

The CLASSY program was developed to fit mixtures of multivariate normal distributions to multichannel, multiacquisition spectral data sets. It thus serves simultaneously as a density estimator (providing an unconditional likelihood for an observation) and a clustering algorithm (providing a conditional likelihood). Since it is anticipated that a particular segment will exhibit characteristic features in a sequence of Landsat acquisitions, it is further anticipated that CLASSY clusters will describe these features.

A primary goal of the developers of the CLASSY program was to assist in the estimation of crop proportions in a Large Area Crop Inventory Experiment (LACIE) segment. This may be accomplished if the proportion of each crop of interest in a cluster can be estimated. If a small subset of the pixels can be labeled using some other procedure, the resulting training set will allow the estimation of the proportions by the maximum likelihood method. Such a procedure was developed on the Laboratory for Applications of Remote Sensing (LARS) system at Purdue University (ref. 1). As a first test, a training set of approximately 100 pixels for each of 10 segments was labeled from ground-truth data. The estimated proportion of small grains was compared to the proportion estimated from ground truth, with very encouraging results.

For the present study, the training set was labeled by an analyst/interpreter (AI), and approximately 200 dots per segment were used. Thus, the experiment more nearly approximates applications conditions. Ten LACIE Transition Year segments were analyzed using the maximum likelihood labeling technique, and, using two different estimates, the proportion of small grains was compared to ground truth, to the standard Procedure 1 (P1) estimate, and to an estimate from a linear classifier trained on labeled dots. This linear classifier was part of the Statistical Analysis System (SAS) package (ref. 2).

2. DATA

Ten LACIE Transition Year segments were chosen under the constraints (1) that there should be some geographical variety, (2) that spectral data would be available at LARS, (3) that four acquisitions of good quality would be available, and (4) that approximately 200 AI-labeled dots would be available for each acquisition. In all cases, the statistic of interest was the proportion of small grains grown in that year. The training dots had previously been labeled by AI's for the P1 crop proportion estimates. As a result, P1 estimates and ground-truth values of the statistic in question were available for comparison. For each of the acquisitions, the spectral data were first projected onto the Kauth-Thomas greenness-brightness plane (ref. 3), and all analyses were conducted on this reduced data set.

3. MAXIMUM LIKELIHOOD PROPORTION ESTIMATION AND CLUSTER LABELING PROCEDURE

The following is a description of the maximum likelihood proportion estimation and cluster labeling procedure evaluated in this study. The purpose is to obtain estimates of the proportion of the class of interest (in this experiment, small grains) in each component distribution or cluster generated by the CLASSY program.

Suppose that the CLASSY program is used to approximate the multivariate mixture density of the data. This will result in a set of multivariate normal distributions $p(x|i)$, $i = 1, \cdots, c$, and a set of prior probabilities α_i, $i = i, \cdots, c$. Now suppose that we have a set of data points x_j, $j = 1, \cdots, N$, and a set of possible class labels ϕ_ℓ, $\ell = 1, \cdots, M$. Then, the joint probability of observing data point x_j associated with label ϕ_ℓ may be formulated as follows.

$$\begin{aligned} p(x_j,\phi_\ell) &= \sum_{i=1}^{c} \alpha_i p(x_j,\phi_\ell|i) \\ &= \sum_{i=1}^{c} \alpha_i p(\phi_\ell|x_j,i)p(x_j|i) \qquad (1) \end{aligned}$$

Assume that

$$p(\phi_\ell | x_j, i) = p(\phi_\ell | i) = \beta_{\ell i} \tag{2}$$

which implies that the label random variable ϕ_ℓ is conditionally independent of the observation x_j; i.e., given that one is sampling from distribution i, no further information is conveyed by knowing x_j.

Using this model, we see that the proportion of class ℓ may be estimated as

$$p(\phi_\ell) = \sum_{i=1}^{c} \alpha_i \beta_{\ell i} \tag{3}$$

and $\beta_{\ell i}$ may be interpreted as the proportion of distribution i that is composed of class ϕ_ℓ.

Alternatively, each cluster may be labeled by selecting the class with the largest value of $\beta_{\ell i}$. A proportion estimate may then be obtained as follows:

$$P(\phi_\ell) = \sum_i \alpha_i$$

$$\text{for all } i \text{ such that } \beta_{\ell i} = \max_j \beta_{ji} \tag{4}$$

The first proportion estimator will be called a stratified maximum likelihood estimator, and the second will be called a labeled cluster maximum likelihood proportion estimator.

To estimate $\beta_{\ell i}$, a maximum likelihood approach may be used, assuming that all α_i and $p(x_j | i)$ are known.

Given a random sample of labeled data points, the likelihood function is

$$L = \prod_{\ell=1}^{M} \prod_{j_\ell=1}^{N_\ell} p(x_{j_\ell}, \phi_\ell) \tag{5}$$

where x_{j_ℓ}, j = 1, $\cdots$, N_ℓ, are those data points labeled as coming from class ϕ_ℓ.

Under the model, the likelihood function may be written as

$$L = \prod_{\ell=1}^{M} \prod_{j_\ell=1}^{N_\ell} \sum_{i=1}^{c} \alpha_i \beta_{\ell i} p(x_{j_\ell} | i) \tag{6}$$

Taking the log of the likelihood function and introducing the constraint that $\sum_{\ell=1}^{M} \beta_{\ell i} = 1$ for $i = 1, \cdots, c$, using Lagrangian multipliers, the function to be maximized becomes

$$F = \sum_{\ell=1}^{M} \sum_{j_\ell=1}^{N_\ell} \log\left[\sum_{i=1}^{c} \alpha_i \beta_{\ell i} p(x_{j_\ell} | i)\right] - \sum_{i=1}^{c} n_i \left(\sum_{\ell=1}^{N} \beta_{\ell i} - 1\right) \tag{7}$$

Maximizing with respect to the $\beta_{\ell i}$ results in a solution of $\frac{\partial F}{\partial \beta_{\ell i}} = 0$, which is given by

$$\beta_{\ell i} = \frac{S_{\ell i}}{\sum_{\ell=1}^{M} S_{\ell i}} \tag{8}$$

where

$$S_{\ell i} = \sum_{j_\ell=1}^{N_\ell} \left[\frac{\alpha_i \beta_{\ell i} p(x_{j_\ell} | i)}{\sum_{i=1}^{c} \alpha_i \beta_{\ell i} p(x_{j_\ell} | i)}\right] \tag{9}$$

Thus, the $\beta_{\ell i}$ terms may be estimated using a fixed-point iteration scheme beginning with

$$\beta_{\ell i} = \frac{1}{M} \;\; ; \;\; \ell = 1, \cdots, M \text{ and } i = i, \cdots, c \tag{10}$$

4. RESULTS

The results of the experiment are given in table 1.

a. Column 1 contains the LACIE segment number.

b. Column 2 contains the percentage of small grains found in the segment, as computed from ground-truth information.

c. Column 3 contains an estimate of the small-grain percentage obtained by clustering every other pixel in the segment using the CLASSY program. The maximum likelihood procedure described in the previous section was then used to estimate the proportion of small grains in each CLASSY cluster, using a set of approximately 200 AI labels for training. A weighted average small-grain percentage estimate was then attained by multiplying the size of each cluster (i.e., its prior probability) by its proportion of small grains and summing over clusters. See equation (3).

d. Column 4 contains the proportion of the AI labels that were labeled small grains; i.e., a simple random sample of the approximately 200 training pixels.

e. Column 5 contains an estimate of the proportion of small grains obtained by calling an entire CLASSY cluster small grains or nonsmall grains, if it had been estimated to contain a plurality of one category or the other using the maximum likelihood technique. Thus, the estimate is the sum of the prior probabilities of the small-grain clusters. See equation (4).

f. Column 6 is the P1 estimate of small grains actually obtained during the LACIE Transition Year processing of each segment.

g. Column 7 is an estimate of the percentage of small grains obtained by deriving a Fisher linear classifier (using equal a priori probabilities) from the labeled training pixels. This classifier was then applied to every second pixel in the segment, and the proportion of those classified as small grains was computed. Thus, exactly the same data were used for this estimate as were used for the maximum likelihood cluster labeling procedure.

TABLE 1.- SUMMARY OF PROPORTION ESTIMATES

Segment (1)	Ground truth (2)	Stratified maximum likelihood, % (3)	Simple random sample, % (4)	Maximum likelihood cluster labeling, % (5)	P1, % (6)	Fisher linear classifier (equal a priori probabilities), % (7)	Fisher empirical Bayes linear classifier, % (8)
1394	35.45	35.64	33.49	27.66	40.0	34.67	28.53
1457	47.72	31.20	30.35	25.71	31.0	36.41	27.59
1518	34.16	26.81	20.63	20.67	28.0	32.78	25.93
1584	51.58	45.38	45.93	36.93	55.0	49.10	48.35
1602	30.42	24.41	23.53	21.79	22.8	31.84	22.22
1619	47.91	38.32	36.68	39.19	50.4	47.61	37.33
1668	9.49	7.49	7.55	6.34	5.0	11.77	6.93
1825	26.69	22.75	22.01	19.33	29.5	28.95	22.71
1909	22.35	10.18	10.05	9.34	9.0	23.61	6.26
1918	15.02	18.54	17.31	18.16	21.0	25.94	12.65
Mean bias		-6.01	-7.33	-9.57	-2.94	.19	-8.22
Mean squared error		67.4	87.6	134.0	65.1	27.0	99.7

h. Column 8 is an estimate of the percentage of small grains obtained from the Fisher empirical Bayes linear classifier, where the crop a priori probabilities are assumed equal to the simple random sample estimate obtained from the AI dots.

At the bottom of each column of estimates is the mean error in the percentage of small grains as observed from ground truth (the bias) and the mean squared error in the estimated percentage of small grains.

5. CONCLUSION AND RECOMMENDATIONS

The results presented in table 1 suggest that the Fisher linear classifier with equal a priori probabilities (column 7) was the best proportion estimation technique both in terms of bias and mean squared error. There seems to be no reason to expect this advantage from the linear classifier. Two possible interpretations are presented: (1) Perhaps the linear classifier requires relatively fewer parameters to be estimated than either the maximum likelihood techniques or P1; this may have led to a stability which was expressed in the lower mean squared error. (2) Although AI labels introduced biases into the proportion estimates, these same labels were, on the average, fairly good at characterizing the distribution of small-grain and nonsmall-grain signatures.

Thus, the Fisher classifier with equal a priori probabilities was able to effectively use the AI's abilities to characterize the signature of small grains and nonsmall grains. This conjecture is supported by the results obtained from the Fisher empirical Bayes linear classifier (column 8). This classifier is identical to the Fisher classifier used to obtain the results shown in column 7, except that the a priori probabilities of the crops were set equal to the simple random sample estimate obtained from the AI dots used to train the classifier. The bias and mean square error obtained for the Fisher empirical Bayes linear classifier are very similar to the results obtained from most of the other proportion estimation procedures.

The stratified maximum likelihood estimate and P1 performed about equally in terms of mean squared error. This result is contrary to previous studies, which indicated a reduction in mean squared error for the stratified maximum likelihood estimation technique as compared to P1 when both techniques used about 100 ground-truth-labeled pixels. It appears that errors present in the AI labels are sufficient to nullify any advantage of stratified maximum likelihood estimation over P1.

Stratified maximum likelihood estimation exhibits an advantage over the simple random sampling method. This result is consistent with the previous results obtained using ground truth.

Cluster labeling using maximum likelihood estimates of cluster purities was the least accurate procedure. This technique performed poorly because of the fact that clusters are not always pure and also because erroneous labels are often sufficient to change the label of a cluster from that which would have been obtained using ground truth.

It is recommended that the maximum likelihood approach to proportion estimation using CLASSY cluster statistics be studied further, both to understand its potential and to pave the way for related but improved techniques for incorporating label information (particularly AI label information) into the process of cluster proportion estimation and labeling. The significance of the essentially unbiased result obtained with the Fisher linear classifier should be investigated. This finding may have implications concerning the fundamental process of AI labeling.

6. REFERENCES

1. Horton, C.; and Lennington, R. K.: "As-Built" Design Specification for CLASSY and Adaptive Maximum Likelihood Clustering Method. JSC-16466, LEMSCO-14546, NASA/JSC (Houston), Feb. 1980.

2. Barr, A. J., et al: A User's Guide to SAS76. The SAS Institute (Raleigh, N.C.), 1976.

3. Kauth, R. J.; and Thomas, G. S.: The Tasselled Cap — A Graphic Description of the Spectral Temporal Development of Agricultural Crops as Seen by Landsat. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University (W. Lafayette, Ind.), June 29 - July 1, 1976.

NASA-JSC